

Knowledge Discovery using Patent analysis

Course Title: **Knowledge Management** Course Number: **ETM 567/667** Instructor: **Dr. Charles Weber** Term: **Winter** Year: **2015** Author(s): Monticha Khammuang

ETM OFFICE USE ONLY

Report No.: Type: Student Project Note:

Table of Contents

1. Abstract	
2. Introduction	4
3. Literature Search	5
3.1 Overview of Knowledge discovery process	5
3. 2 Overview of Patent analysis and its application	
3.3 Knowledge discovery process using Patent analysis	10
4. Evidence from past experience	12
5. Discussions & Recommendations	16
6. Conclusions	
7. References	19

1. Abstract

Knowledge Discovery helps companies to understand of how thing work in organizations and in the nature. Its contribution has shaped their performance in many ways such as it could help company to create the new innovative product or could help company to solve the problems. Moreover, companies could use it to forecast the future market trend or use it for future collaboration.

Nowadays, there are many tools that could help companies to discover the knowledge. Patent analysis is one among those tools. This paper examines challenges associated with knowledge discover by using Patent Analysis commonly experience by the author. While others have researched this topic several times in the past, the author studied by observing the past researches that due to what is out there, what is the process of knowledge discovery, the impacts from using Patent Analysis and then compare what are found from the past researches with her past experience. Additionally, the author also gives some recommendations on Knowledge Discovery using Patent Analysis to her organization by using her own learning from both from class and from previous researches.

Keywords: Knowledge Discovery, Patent Analysis, Text Mining, Patent Databases

2. Introduction

The general aim of Knowledge Discovery is to extract unknown and potentially important information from data. According to the rapid growth of scientific, business and industry databases, many knowledge discovery techniques have been used regularly. In addition, as the usual Data mining techniques are necessarily designed to operate on structured databases, specific techniques like Text mining techniques have been improved to operate the available information that could be found in unstructured data. Text mining therefore corresponds to the extension of the more traditional Data mining approach to unstructured textual data and is primarily concerned with the extraction of information implicitly contained in collections of documents [1].

Patent databases combine of text data that explain a patent technology and its applications. The original method of extracting knowledge from patent databases was based on manual analysis operated by experts in that field. Nowadays this method is impractical because patent databases grow so fast. In addition, more and more R&D planners, business analysts, patent analysts, national or international patent offices, economic organizations, national statistical offices and industrial parts are interested in knowledge discovery and exploitation from useful data hidden in patent databases [2].

In this paper, the author presents the Knowledge Discovery process, Knowledge discovery meaning, Patent Analysis process and applications, and the role of Patent Analysis on Knowledge Discovery in section 3. Then, the example from her past experience on Knowledge Discovery process using Patent Analysis is reviewed in section 4. The discussions and recommendations are shown in section 5 and finally the conclusions are given in section 6.

4

3. Literature Search

3.1 Overview of Knowledge discovery process

Knowledge discovery is a process of discovering useful patterns from the original large data sets. The Knowledge discovery process is interactive and iterative involving several steps with many decisions making from the user. Almost of the existing methods for knowledge discovery show the interactions between the human and the machine. These interactions help Knowledge discovery drive through the gigantic search spaces and acknowledge the intentions of the user, and the user feeds the system easily with the core knowledge and factors. In contrast, the visualization of the Knowledge discovery gives better insight about the multidimensional data to users. They could understand the results and could interpret the discovery patterns well [3].

Although the Knowledge discovery process can be widely separated into two groups: the industrial approach and academic approach.

• Knowledge discovery for Academic purposes [3]:

There are nine steps of Knowledge discovery for Academic purposes. Starting with, user should develop and understand the application domain including learning the related prior knowledge and the aims of the users of the discovered knowledge. Then users have to create a target data set, variables, and data points. After that they have to remove the outliers, delete the noise and deal with missing values in the data. Then move to the next step, which is the data reduction and projection. In this step, they have to find useful attributes by applying dimension reduction and trying to find invariant representation of the data set. Next, they have to choose the data-mining task by matching the goals defined in step 1 with a particular Knowledge discovery method such as patenting, clustering, regression, or classification. After that they have to start mining or finding the patterns from the data set and interpret the mined patterns. Finally, analysts merge the discovered knowledge into performance system, document and report it to the interest groups.

• Knowledge discovery for Industrial purposes [4]:



Figure 1 shows the CRISP-DM KDP model (Source: <u>www.crisp-dm.org</u>)

The above figure shows the CRISP-DM KDP model. This model represents six steps of Knowledge discovery process for Industrial purposes.

➤ Step 1: Business Understanding. This step point on the understanding of objectives and requirements from a business perspective. It also changes these into a Knowledge discovery problem statement, and designs a project plan to accomplish the objectives. It combines of determination of business objectives, an assessment of the situation, determination of Knowledge discovery goals and generation of a project plan.

Step 2: Data Understanding. In this step, users have to collect initial data and get used to with the data including identification of data quality problems, initial insight into the data and detection of interesting data sets.

Step 3: Data Preparation. This step covers selection of data cleansing of data, construction of data, integration of data, and formatting of data subsets.

Step 4: Modeling. At this step, users have to select modeling techniques that they to use in this step. Then they need to generate the test design, create the models and assess the model generated by the users.

➤ Step 5: Evaluation. After models have been built, they need to evaluate the results and review the process. The main issue is to determine whether any important business issue has not been sufficiently considered. Then determining the next step.

Step 6: Deployment. Now the discovered knowledge must be structured and showed in a way that users/customers can use. The final report need to be made as well as the review of the process sub-steps.

From both knowledge discovery processed for Academic and Industrial purposes, we could see that the knowledge discovery process consists of multiple steps that are structured in sequence. The subsequent step is initiated upon successful completion if the previous step and requires results generated by the previous step as its inputs. The knowledge discovery process varies from understanding the project domain and data, through data preparation and analysis, to evaluation, understanding and application of the generated knowledge. It is highly iterative and includes many feedback cycles and responds, which are irritated by revision processes.

7

3. 2 Overview of Patent analysis and its application

Patent analysis is a methodology used to extract the hidden information in patents, which are stored in international databases. Patent analysis can help experts in intellectual property focus their time and efforts on those patent and activities that add the most benefit to organization. Moreover, Patent analysis also has application to many different tactical and strategic organization evaluations. Patent analysis can provide a very clear perspective of the current trends in terms of technological scientific innovation but in order to exploit this hidden information, it is very important for companies to understand the methodology used in Patent analysis as well as understand the its application [1].

According to [5], there are six main purposes of using Patent analysis in industry.

> 1^{st} purpose: Patent inventory. Companies usually have lists of its patents. They usually store their patents in a Patent inventory. Patent inventory is like a warehouse used to store and maintain a complete and up-to-date database of its patents. Moreover, companies also search public patent databases to find out what the public record says and bring them to their Patent inventory. It seems easy but it isn't.

 $> 2^{nd}$ purpose: Technology transfer and licensing. Technology transfer and licensing is another important application of patent such as patent analysis can help to identify specific patents that may not be current value but may be of interest to other companies so they could offer licensing to companies most likely to be interested in those patents. Moreover, some companies use Patent analysis as a source of knowledge transfer for both internally and externally.

 $> 3^{rd}$ purpose: Donation targeting. Companies normally have patents that they have no commercial use. One way to gain benefit from these is to give them to colleges,

universities or other public organization and receive a tax reduction. The amount of tax deduction depends on the ability to argue that the patent has value to others.

 $> 4^{th}$ purpose: R&D management/technology assessment. Patents can also use as indicators or indirect evaluators of company commitment to R&D and technology. If one company has more patents than another, this suggests that the first company has a stronger commitment to R&D. Moreover, it shows that there is a possibility that the first firm is able to develop new products, processes, or services based on this technology better than others. In addition, we also could know who is the leader in the industry from the number of patents one company invented. Normally, Company who owns the highest number of patents in that technology is likely to be the leader in that technology.

 $> 5^{\text{th}}$ purpose: Forecasting trends of technology. Company could use the Patent analysis in order to monitor the future trends of technology by observing the most cloudy patents area. The area of patent that contain many patents means that the technology in that field is being interested by many groups of companies and they are possible to become lead technology in the future.

 $> 6^{th}$ purpose: Knowledge discovery source. Patent contains useful information and Patent analysis is a tool that could help company to convert from information to knowledge. Many companies are now interesting in this purpose because they want to convert the hidden information in patents to new products/ processes/ services that could benefit their organization. This purpose becomes critical issue. Many companies give an important to this purpose because they know that if they could discover the knowledge from patent application, they might yield a wealth of information related to research activity such as the key to new invention, the keys to new technology, collaborations, location of research work, key inventors and licensing.

3.3 Knowledge discovery process using Patent analysis

The process of knowledge discovery using Patent analysis is shown in figure 2.



Figure 2 shows the process of Knowledge discovery using patent analysis [6]

From figure 2, there are three main steps to discover knowledge by using patent analysis [6].

Step 1: Textual Analysis. Analysts need to read the patents data from the given database and change them into the proper format for the next processing.

➤ Step 2: Linguistic Preprocessing. This step is to understand the meaning of texts. Its main goal is to change the raw data of patents into a proper format in order to use for future statistical analysis. The activities in this step are include of Data cleansing – analysts need to clean input data by removing irrelevant characters, Lemmatization –

analysts need to decrease each of the different inflections to a unique representation, Part of speech tagging – analysts need to identify the categories of words.

Step 3: Clustering Process. This process start by giving each point to each own cluster and compute the distance between the points. Next find the closet pair of points and combine them into a single cluster. Each new cluster represents a node in the produced hierarchical tree. The above procedure is cycled until analysts regroup all the points into one single cluster.

➤ Step 4: Correspondence Analysis. This step is to explore the cross-classified data by searching low dimensional geometrical representations and related numerical statistic.

➤ Step 5: Visualization. This step is used for the results' extraction and the visualization. There are many options at this step such as graph, table, or picture. This step is very important because it is the step that analysts explain what they found to other people and the patterns they found also represent the knowledge discovered from patents.

4. Evidence from past experience

5 years ago I worked as production engineer in The Siam Molding Plaster company (It is part of Siam Cement Conglomerate). My main job was to taking care of the production in Grinding and calcination of Calcium Sulfate process. Moreover, because I have background in R&D job so I was assigned additional job to take care about patent analysis. The main tasks of Patent analyst could be seen in figure 3 [7].

Figure 3: The main tasks of Patent Analyst



Source: Siam Molding Plaster Company [7] (Please do not distribute or publish)

From Figure 3, at that time there were three main tasks of Patent Analyst. First was to filing a new patent or IP. If there was a new patent in company in any department,

patent analyst had to talk with that department and help them to write and apply for the patent. Secondly, Collaboration with external parties such as SCG IP management center for strategic planning or with ministry of Intellectual Property of the Thai government for seeing the published filed patent. Lastly about Knowledge discovery, this task was the main job of patent analyst. There were IP committees in company. Patent Analyst has to report all progress about patent to committees. Moreover, patent analyst is in charge for searching the patent that might create value to company and to monitor competitors' movement in patent as well as forecast the technology trends. The processes of knowledge discovery in company could be seen and figure 4.

Figure 4: The processes of knowledge discovery in company



Source: Siam Molding Plaster Company [7] (Please do not distribute or publish)

From figure 4, the processes of Knowledge discovery is starting from Patent analyst search for patents by using Aureka databases program. Our SCG conglomerate bought the program and allowed companies under conglomerate to use it. Analyst will search with different keywords. After analyst got patent information, she will look through each patent roughly and then assign to all engineers in company to read and study patents (this company is stall company with about 150 employees, 15 engineers including top management). Next, all assigned engineers have to present the detail of patents to committees. They can use any kind of presentation tool such as power point, word or excel. Then IP committees will consider if they see the value of that patent, they will assign team to study it more deeply but if there is no value or not yet give the value at that time, the patent will be kept in patent warehouse. The IP committees also monitor the progress of the project after assigned by following up in R&D meeting. The result searched from using Aurora databases could be seen in figure 5 and 6 respectively.

Figure 5: Result from using Aureka database

Aureka - Microsoft Internet Explorer					
File Edit View Favorites Tools Help			in Ly, i y		🙀
🌀 Back 🔹 🕥 - 💌 🖻 🏠 🏓	🔿 Search 👷 Favorites) 🗟 • 🍓 🖻 • 🗔 1	8		
Address Address http://aureka.micropat.com/7w/htm	l/aos_app_main.asp				Go Links »
aureka®	Enter Patent Number (e.g., <u>US555</u>	80420_EP0795855_WO9725803) र Go			Home Support Log out
Gold	👻 Search 😭 Citations 🎲	Alerts 👔 Reports 👩 Maps 👔	Documents 🌇 Prefere	ences	
0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	Document List 🗈	Action Edit Application Vi	ew Help		
E-@TCERAMIC	1 Patent V Documents of	of 1 Total Documents			Print-Friendly
Apr 27,07-List	Q P Results 1 - 1	No documents sele	ted.		Page 1 of 1
C04B Apr 27,07-List (DeDup		Title Therapeutic agent for steroid depend	Assignee	Inventor Obkusa Toshifumi	Pub Date App Date
C04B Apr 27,07-List (E	A D Reculte 1 1	merapeute agent for ateroid depent	ang-ana Asirono ro rac	onkasa, rosiniani	Door 1
Dune 29, 07	V V Results 1-1				Page of 1
I tablet plaster-Query tablet plaster-Query tablet plaster-List wall plaster Lafarge on Feb 09,07 Automobile plaster-List metal casting-List (Del metal casting-List (Del matal casting-List (Del matal casting-List (Del metal casting-List		Terms and Conditions	Erivaor Polior	E Thomson 2008	
<					
Done					Internet
🛃 Start 🛛 🗿 Aureka - Microsoft In	. 🔯 http://aureka.microp	C:\Documents and Se		EN S	🕈 🗊 🗑 関 🧐 🏷 9:55 AM

http://aureka.mic	ropat.c	om - US20060	106002	A1 Aureka Online	- Microsoft In	ternet Explo	orer							B X
File Edit View Fave	orites	Tools Help												
U\$200601060	02A1	Action Edit	Help											
Summary Text	Claim	s Image	PDF	Legal Status Lir	ked Notes									
cultury														
													Print-Frie	endly
Title:		litle:	-	herapeutic agent to	r steroid depend	ding-and/or st	teroid resistant	-ulcerative colitis	;	id denou	a dia a a	a diar ata	raid raai	ative
		L F	The present memory is directed to a pharmaceutical composition for the realment of send depending and/or service ulcerative colitis. The pharmaceutical composition, which is administered to treat steroid depending and/or steroid resist ulcerative colitis, contains, as the active ingredients, at least one compound selected from a group consisting of tetracycli penicillins, and nitroimidazoles.						resistive racycline	ve nes,				
		Assignee:		ЈІЛОМОТО КК										
	Inventor:			Ohkusa, Toshifumi Batoh, Nobuhiro										
		Publication D	ate: 2	006-05-18										
		Application D	ate: 2	005-09-14										
		Cites:	()										
		Cited By:	(
	Intl Class:				Core: A61K00314164 [2006-01]; A61K0031429 [2006-01]; A61K003165 [2006-01] Adv: A61K00314164 [2006-01]; A61K003143 [2006-01]; A61K003165 [2006-01]									
		ECLA:												
		US Class:	1	14152; 514192; 51	4398									
		Field of Searc	ch:											
Done												Interne	t	
📲 start 🛛 🗿	Aureka -	Microsoft In	a be	n://aureka micron	Microsoft	PowerPoint				EN	0 2 mil		0.8. 0	57 AM

Figure 6: Result from using Aureka database

5. Discussions & Recommendations

If we compare figure 4 (process of knowledge discovery from Siam Molding Company) with figure 1 (the CRISP-DM Knowledge Discovery Process model) of literature part, we could see the three differences.

First point is about Business understanding step. The literature [4] mentioned that before set up the analysis or the project company needs to focus heavily on the objectives and requirements from business perspective but this step is missing in figure 4. Siam molding company could add this step in there patent process in order to understand more clear about what they want so they could plan about patent in right direction and does not waste time on the project or patent that does not create much value to company.

Second point is about Data preparation step. The literature [4] explained that analysts have to prepare patent data well. Data should be construct in proper format and if analysts want to analyze, the data should be group in the same structure and clean them before analysis. Although Siam molding company did well in searching patents and distribute to their engineer but the data was roughly clean. This might make their engineers waste their time from reading and find that the assigned patent is not worth for their company. Moreover, the patents searched by analyst were rarely group and structure information so company should consider this point because this information might give some useful information such as the location of lead technology, the source of collaboration, technology trends, and market opportunity.

Final point is about modeling step. The literature [4] said that modeling technique could be implemented in knowledge discovery in order to find the patterns of the information. Nowadays, the raw material of Siam molding Company is from natural

16

resources. One day, this raw material might gone from Thailand so if company could add this step in their process to find the patterns of their patent information. May be company could able to find the new pattern that could lead to new raw material or new technology that could replace the old technology.

If we compare figure 4 (process of knowledge discovery from Siam Molding Company) with and figure 2 (the process of Knowledge discovery using patent analysis) of literature part, *we could see the one difference, which is about clustering process.* The literature [5] explained that clustering is used to detect natural grouping in data but there is no this step in figure 4 so may be in searching step of Siam Molding Company, they should consider this step and analyze data by using a group of data, not only one data. Group of data could help company to notice the core technology or the trend of technology that might come to the industry in the future.

From author point of view, Siam Molding Company could change their knowledge discovery process using patent analysis as follow.



Figure 7: Knowledge Discovery Process purposed by the author

6. Conclusions

Patents are overall good things to the business sector, industry and society, especially to the high technology companies. They are one of the useful knowledge sources. Organization that could utilize the use of patents, could gain many advantages from it such as they can exploit the patent to create new product, they could monitor competitor, they could found the leader of the technology, they could notice where the collaboration is, or they could found the new solution for their company.

Knowledge Discovery plays important role in current world. It helps people to gain the new pattern of information or knowledge. People could use this new knowledge to benefit their organization or society. The processed are includes of textual analysis or data preparation, clustering process, correspondence analysis or the statistical analysis and the visualization.

The findings from literatures could give some suggestions about Knowledge Discovery process to Siam Molding Company. The author believes that these suggestions could help them to maximize the use of patents in many ways.

7. References

[1] Sirmakessis, Spiros, STING: Evaluation of Scientific & Technological Innovation and Progress in Europe Through Patents, 1st STING User-Focus Group Meeting, Lausanne, 2001, pp. 33 - 44.

[2] Grandjean, Nicolas, Competitive intelligence and patent analysis in drug discovery: Mining the competitive knowledge bases and patents, Drug Discovery Today: Technologies 2.3, 2005, pp. 211-215.

[3] Han, Jianchao, Nick Cercone, RuleViz: a model for visualizing knowledge discovery process, Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining, ACM, 2000, pp. 244 - 253.

[4] Brachman, R., Anand, T, The process of knowledge discovery in databases: a humancentered approach, 1996, pp. 37–58.

[5] Markellos, Konstantinos, Knowledge discovery in patent databases, Proceedings of the eleventh international conference on Information and knowledge management, 2002, pp. 672 - 674.

[6] Cabena, P., Hadjinian, P., Stadler, R., Verhees, J., Zanasi, A Discovering Data Mining: From Concepts to Implementation, Prentice Hall Saddle River, New Jersey, 1998, pp. 187 – 205.

[7] Patent report in company meeting, The Siam Molding Company, July 2009.